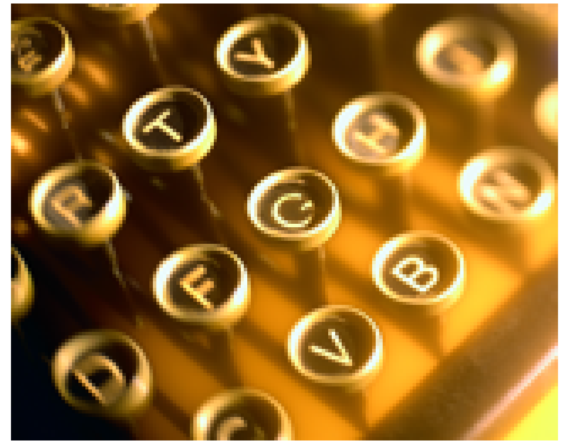
Stability and Selection in Game Theoretic Learning

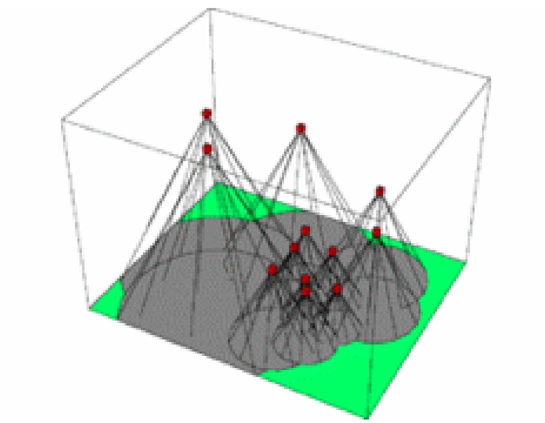
Jeff S Shamma
Georgia Institute of Technology

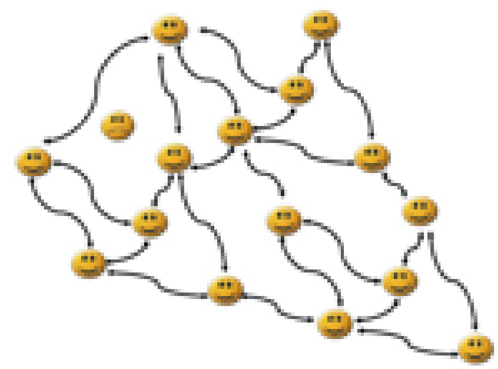
Joint work with
Gürdal Arslan, Georgios Chasparis & Michael J. Fox

Valuetools 2011
Georgia Institute of Technology
May 18, 2011

Networked interaction: Societal, engineered, & hybrid







- Game elements:
  - Actors/players
  - Choices
  - Preferences over *collective* choices
  - Solution concept (e.g., Nash equilibrium)
- Descriptive agenda:
  - Modeling of natural systems
  - Game elements *inherited*
  - Modeling metrics
- Prescriptive agenda:
  - Distributed optimization for engineered (programmable!) systems
  - Game elements *designed*
  - Performance metrics

Arrow, 1987: *The attainment of equilibrium requires a disequilibrium process.*

Skyrms, 1992: *The explanatory significance of the equilibrium concept depends on the underlying dynamics.*

Background: Game theoretic learning

Arrow: "The attainment of equilibrium requires a disequilibrium process."

Skyrms: "The explanatory significance of the equilibrium concept depends on the underlying dynamics."

- Monographs:
  - Weibull, *Evolutionary Game Theory*, 1997.
  - Young, *Individual Strategy and Social Structure*, 1998.
  - Fudenberg & Levine, *The Theory of Learning in Games*, 1998.
  - Samuelson, *Evolutionary Games and Equilibrium Selection*, 1998.
  - Young, *Strategic Learning and Its Limits*, 2004.
  - Sandholm, *Population Dynamics and Evolutionary Games*, 2010.
- Surveys:
  - Hart, "Adaptive heuristics", *Econometrica*, 2005.
  - Fudenberg & Levine, "Learning and equilibrium", *Annual Review of Economics*, 2009.

- Single agent adaptation:
  - Stationary environment
  - Asymptotic guarantees
- Multiagent adaptation:

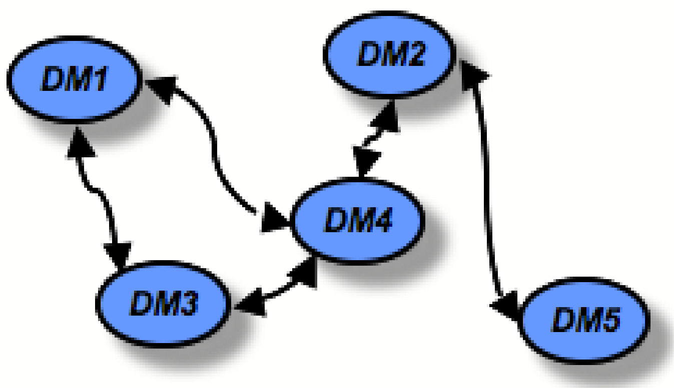
Environment

$=$

Other learning agents

$\Rightarrow$

Non-stationary



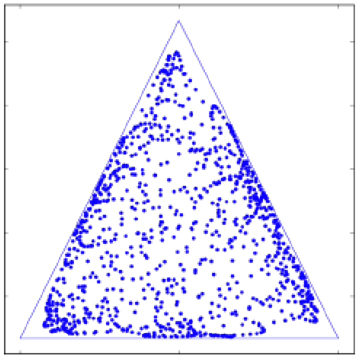
- A is learning about B, whose behavior depends on A, whose behavior depends on B...i.e., **feedback**
- Resulting non-stationarity has major implications on achievable outcomes.

- Setup: Repeated play
- Each player:
  - Maintains empirical frequencies (histograms) of other player actions
  - Forecasts (incorrectly) that others are playing randomly and independently according to empirical frequencies
  - Selects an action that maximizes expected payoff

- Convergence: Zero sum games (1951); 2×2 games (1961); Potential games (1996); $2 \times N$ games (2003).

- Non-convergence: Shapley fashion game (1964); Jordan anti-coordination game (1993); Foster & Young merry-go-round game (1998).

- Setup: Continuous-time "replicator dynamics" on perturbed RPS



- **Sato et al (PNAS 2002):** Chaos in learning a simple two-person game
  "Many economists have noted the lack of any compelling account of how agents might learn to play a Nash equilibrium. Our results strongly reinforce this concern, in a game simple enough for children to play."

Illustration: Stochastic adaptive play & selection

	A	B
A	4,4	0,0
B	0,0	3,3

Typewriter Game

	S	H
S	3/2,3/2	0,1
H	1,0	1,1

Stag Hunt

- How to distinguish equilibria?
- Payoff based distinctions: Payoff dominance vs Risk dominance
- Evolutionary (i.e., *dynamic*) distinction
  - Young (1993) "The evolution of convention"
  - Kandori/Mailath/Rob (1993) "Learning, mutation, and long-run equilibria in games"
  - many more...
- Adaptive play:
  - "Two" players sparsely sample from finite history
  - Players either:
    - ∗ Play best response to selection
    - ∗ Experiment with small probability
  - **Young (1993):** Risk dominance is "stochastically stable"

	Stability	Selection
Descriptive	*explanation*	*refinement*
Prescriptive	*adaptation*	*efficiency*

- Transient phenomena & stability
- Transient phenomena & selection
- Stochastic stability & self-organization
- Network formation, self-assembly, language evolution

- Setup:
  - Players: $\{1, ..., p\}$
  - Actions: $a_i \in \mathcal{A}_i$
  - Action profiles:

$$(a_1, a_2, ..., a_p) \in \mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \times ... \times \mathcal{A}_p$$

  - Payoffs: $u_i : (a_1, a_2, ..., a_p) = (a_i, a_{-i}) \mapsto \mathbf{R}$
- Nash equilibrium: Action profile $a^* \in \mathcal{A}$ is a NE if for all players:

$$u_i(a_1^*, a_2^*, ..., a_p^*) = u_i(a_i^*, a_{-i}^*) \geq u_i(a_i', a_{-i}^*)$$

- Learning dynamics:
  - $t = 0, 1, 2, ...$
  - $\mathbf{Pr}\left[a_i(t)\right] = p_i(t), \quad p_i(t) \in \Delta(\mathcal{A}_i)$
  - $p_i(t) = \mathcal{F}_i($available info at time $t)$

Setup: Continuous vs discrete time dynamics

- Stochastic approximation:

$$x(t+1) = x(t) + \frac{1}{t+1}\Big(\text{rand}[F(x(t))]\Big) \quad \Longrightarrow \quad \frac{dx}{dt} = \overline{F}(x)$$

- Summary: Continuous-time analysis has discrete-time implications
- Illustrations (two player):
  - Smooth fictitious play:

$$f_i(t+1) = f_i(t) + \frac{1}{t+1}\Big(\beta_i(f_{-i}(t)) - f_i(t)\Big)$$

$$\Downarrow$$

$$\frac{df_i}{dt} = -f_i + \beta_i(f_{-i})$$

  - Reinforcement learning:

$$p_i(t+1) = p_i(t) + \frac{1}{t+1} \cdot u_i(a(t)) \cdot \big(a_i(t) - p_i(t)\big)$$

$$\Downarrow$$

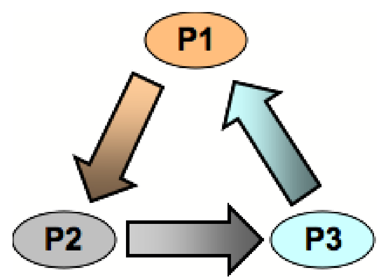
$$\frac{dp_i}{dt} = \Big(\text{diag}[M_i p_{-i}] - \text{diag}[p_i^{\mathsf{T}} M_i p_{-i}]\Big)p_i$$

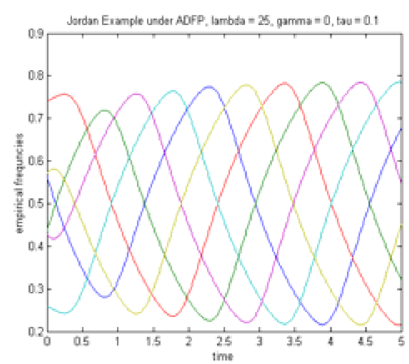
replicator dynamics

- *Uncoupled dynamics:*
  - The learning rule for each player does not depend (explicitly) on the payoff functions of the other players.
  - Satisfied by fictitious play & replicator dynamics

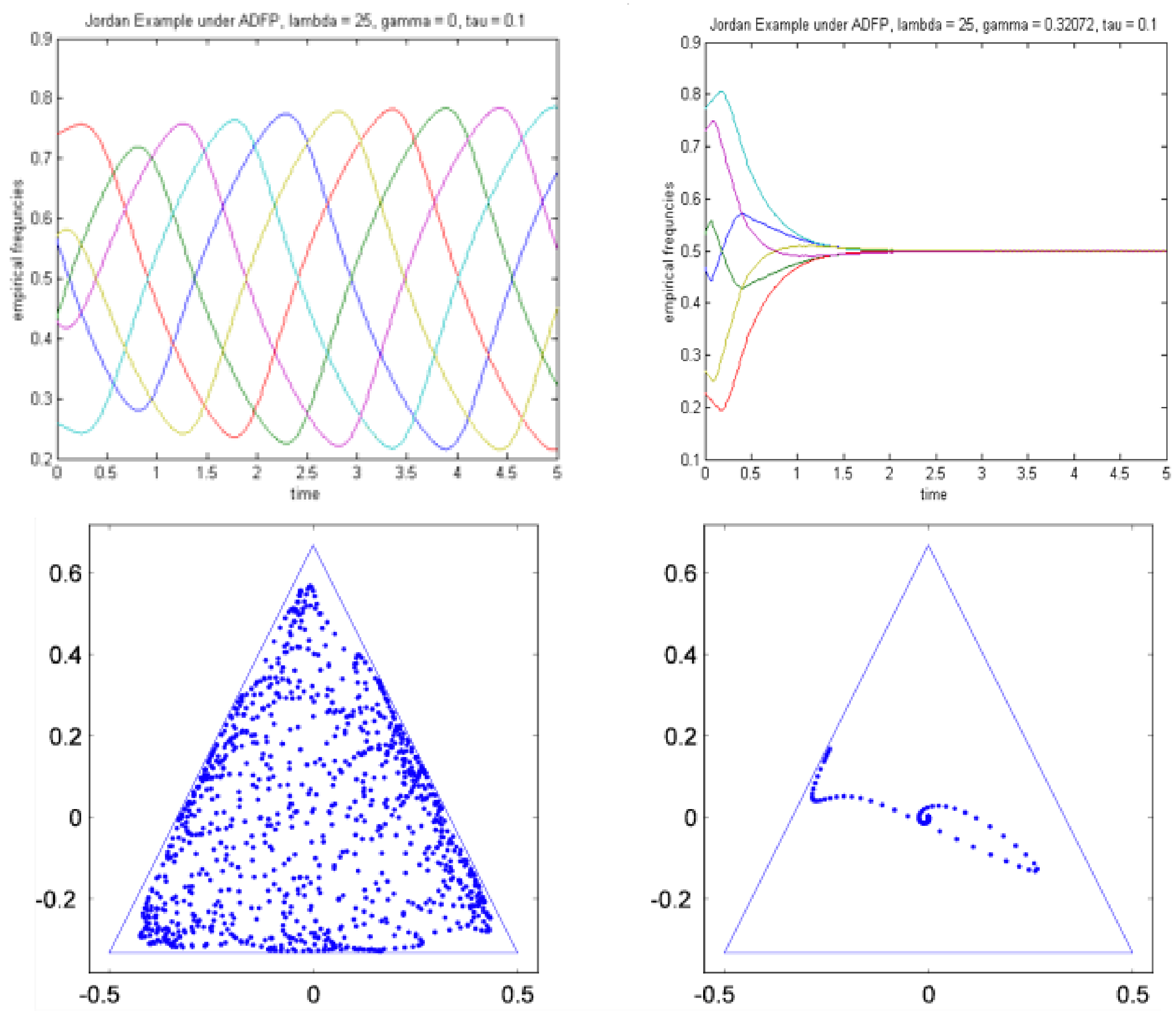
- **Hart & Mas-Colell (2003):** There are no uncoupled dynamics that are guaranteed to converge to Nash equilibrium.
  Analysis: Jordan anti-coordination game is universal counterexample.
  (cf., *Saari & Simon (1978)*)

- Three players & two actions
  - Player 1 $\neq$ Player 2
  - Player 2 $\neq$ Player 3
  - Player 3 $\neq$ Player 1





Uncoupled dynamics & convergence?



- Negative results only apply to *static* learning rules

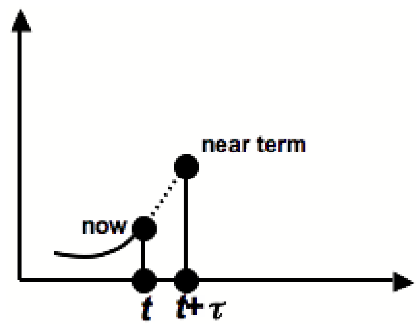
$$\frac{dp_i}{dt}(t) = \overline{F}_i(p_i(t), p_{-i}(t); M_i)$$

(applies to fictitious play & replicator dynamics)

- What about *dynamic* learning rules?

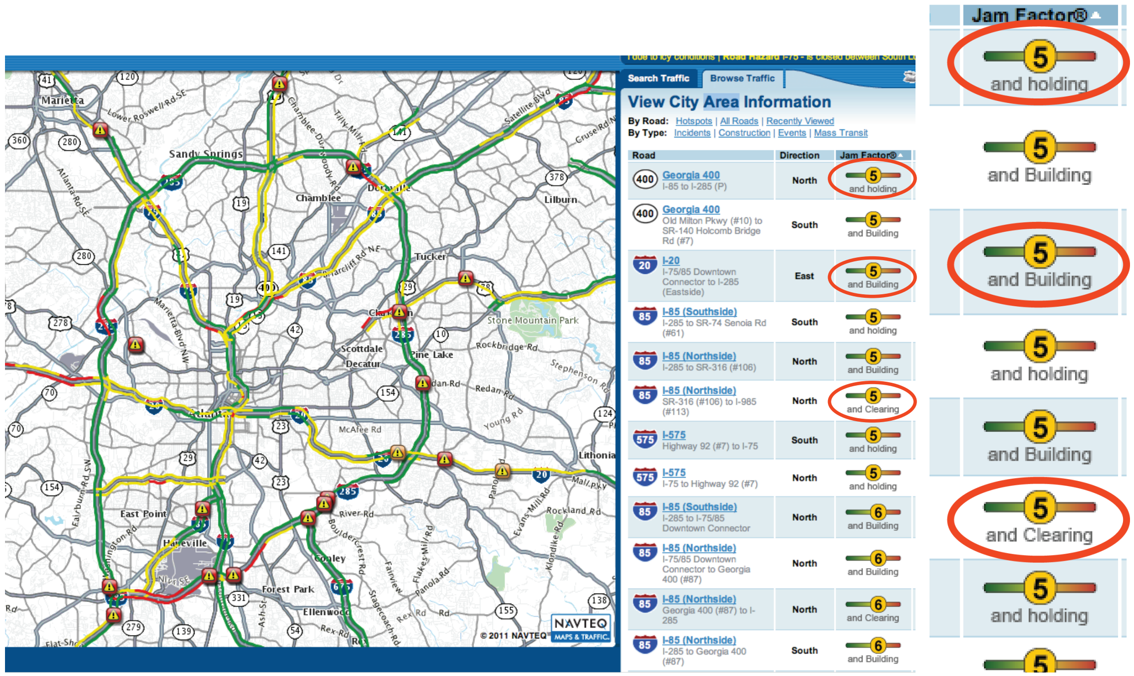
$$\frac{dp_i}{dt}(t) = \overline{F}_i(p_i(\cdot), p_{-i}(\cdot); M_i)$$

- **Marginal forecast dynamics:**
  - React to myopic predictions
  - FP: Best response to forecast empirical frequency
  - Replicator dynamics: React to forecast fitness
- Features:
  - Purely transient
  - Still *uncoupled!*



$$q(t+\gamma) \approx q(t)+\gamma\frac{dq^{\text{est}}}{dt}(t)$$

- ATL traffic: "Jam Factor" *Holding, Building, Clearing*



- Background:
  - Basar (1987), "Relaxation techniques and asynchronous algorithms for online computation of noncooperative equilibria"
  - Selten (1991), "Anticipatory learning in two-person games"
  - Conlisk (1993), "Adaptation in game: Two solutions to the Crawford puzzle"
  - Tang (2001), "Anticipatory learning in two-person games: Some experimental results"
  - Hess & Modjtahedzadeh (1990), "A control theoretic model of driver steering behavior"
  - McRuier (1980), "Human dynamics in man-machine systems"

Analysis: Marginal forecast fictitious play

$$\frac{dr_i}{dt} = \lambda(f_i - r_i)$$
$$\frac{df_i}{dt} = -f_i + \beta_i\Big(f_{-i} + \gamma\frac{dr_{-i}}{dt}\Big)$$

- Approximation for $\lambda \gg 1$:
$$\left|\frac{df_i}{dt} - \frac{dr_i}{dt}\right| \leq \frac{1}{\lambda}\left|\frac{d^2 f_i}{dt^2}\right|_{\text{max}}$$
- *Note:* Auxiliary variables absent from prior impossibility result!
- **JSS & Arslan, 2005:** For large λ
  - FP stable at NE p^* implies marginal foresight FP stable at q^* for $0 \leq \gamma < 1$
  - FP unstable at p^* with eigenvalues $x_k + jy_k$ and
  $$\max_k \frac{x_i}{x_k^2 + y_k^2} < \frac{\gamma}{1-\gamma} < \frac{1}{\max_k x_k}$$
  implies marginal foresight FP stable at p^*.
- Similar results:
  - Marginal foresight replicator dynamics
  - Marginal foresight tatonnement

- Reinforcement learning: $x_i =$ action propensities

$$x_i(t+1) = x_i(t) + \delta(t)(a_i(t) - x_i(t)), \quad \delta(t) = \frac{u_i(a(t))}{t+1}$$

$$p_i(t) = (1-\varepsilon)x_i(t) + \frac{\varepsilon}{N}\mathbf{1}$$

$$\delta_{\text{std}}(t) = \frac{u_i(a(t))}{\mathbf{1}^{\mathrm{T}}U_i(t) + u_i(a(t))}$$

Interpretation: Increased probability of utilized action.

- *Dynamic* reinforcement learning: Introduce running average

$$y_i(t+1) = y_i(t) + \frac{1}{t+1}(x_i(t) - y_i(t))$$

$$p_i(t) = (1-\varepsilon)\Pi_\Delta\left[x_i(t) + \underbrace{\gamma(x_i(t) - y_i(t))}_{\text{new term}}\right] + \frac{\varepsilon}{N}\mathbf{1}$$

- **Chasparis & JSS (2009):** The pure NE a^* has positive probability of convergence iff

$$0 < \gamma_i < \frac{u_i(a_i^*, a_{-i}) - u_i(a_i', a_{-i}^*) + 1}{u_i(a_i', a_{-i}^*)}, \quad \forall a_i' \neq a_i^*$$

  (as opposed to all pure NE)
  Proof: ODE method of stochastic approximation.

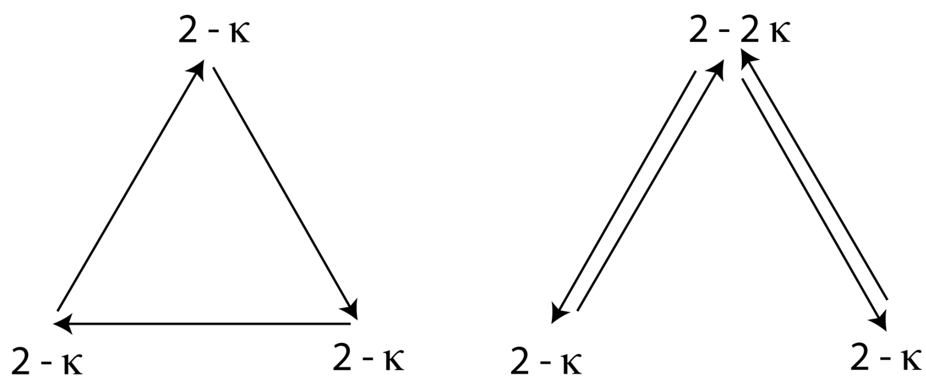
- Implication:
  - Introduction of "forward looking" agent can destabilize equilibria
  - Surviving equilibria = equilibrium selection

- For 2×2 symmetric coordination games
  - RD & not PD $\Rightarrow$ foresight dominance
  - RD & PD & Identical interest $\Rightarrow$ foresight dominance
  - RD & PD together $\not\Rightarrow$ foresight dominance

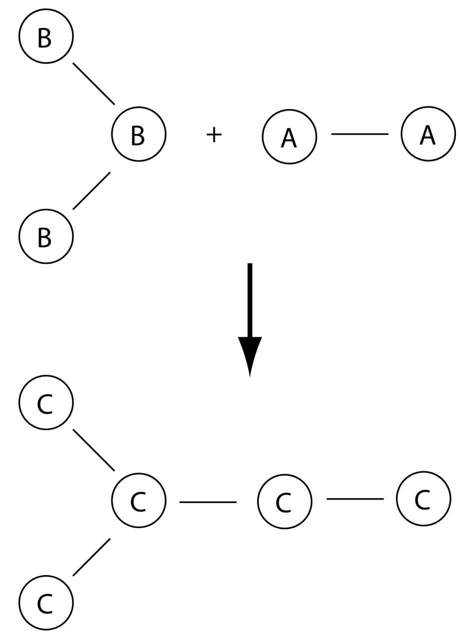
- Setup:
  - Agents form costly links with other agents
  - Benefits inherited from connectivity

$$u_i(a(t)) = \Big(\text{\# of connections to } i\Big) - \kappa \cdot \Big(\text{\# of links by } i\Big)$$

- Properties:
  - Nash networks are "critically connected"
  - Wheel network is unique *efficient* network
  - **Chasparis & JSS (2009):** The wheel network is foresight dominant.

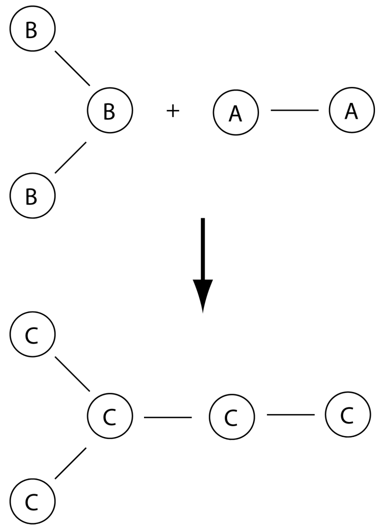


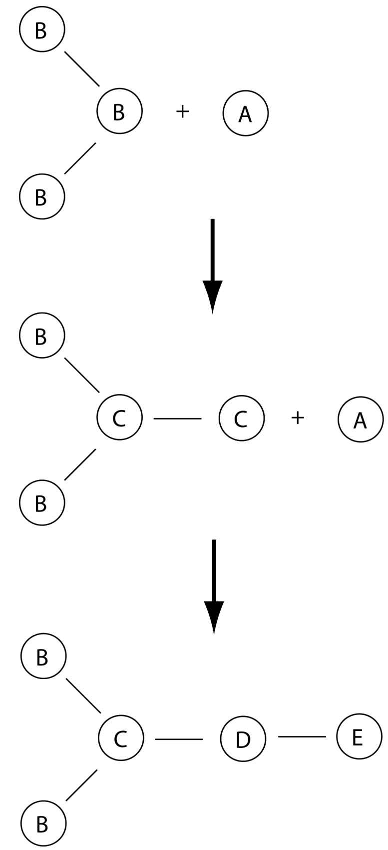
- Recent work considers *transient establishment* costs



- Atoms form subassemblies.
- Subassemblies form complete assemblies.

- References:
  - Yim, Shen, Salemi, Rus, Moll, Lipson, Klavins, & Chirikjian, “Modular self-reconfigurable robot systems: Challenges and opportunities for the future”, 2007.
  - Klavins, “Programmable self-assembly”, 2007.

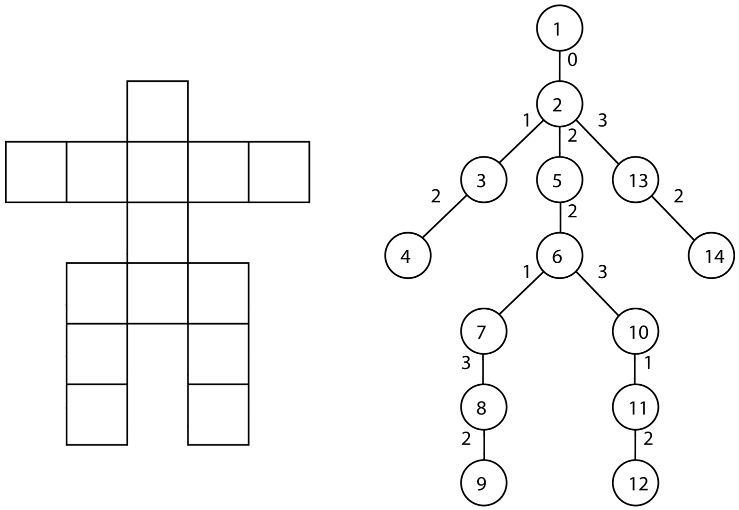




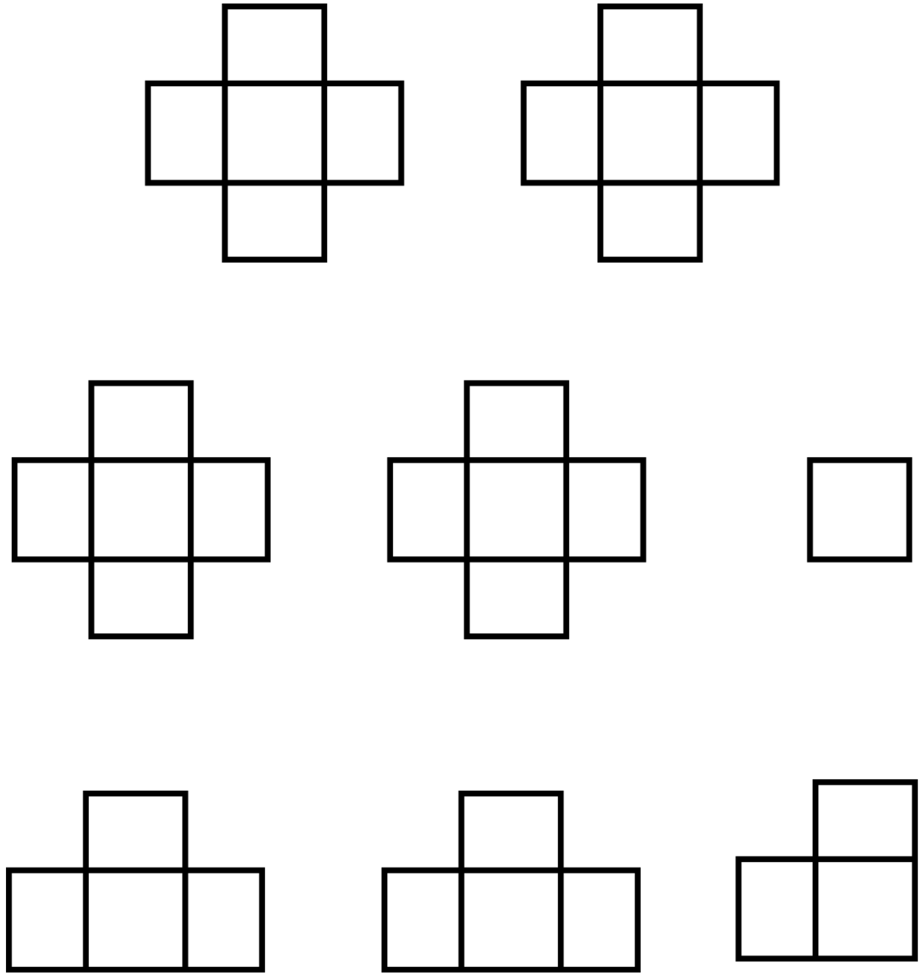
- General setup:
  - Infinite supply
  - Nonlocal rules
  - Full “graph grammars”

- Restricted setup: What is achievable?
  - Finite supply
  - Local rules: Bond or break
  - Reversibility

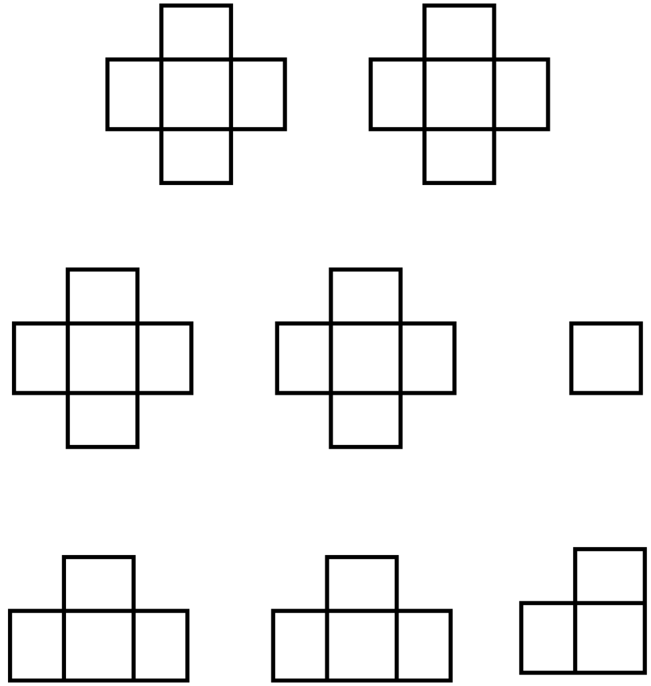
Assembly rules



- Complete assembly = Acyclic weighted graph
- Node state: (Position, Vacancies)
- Nodes meet randomly
- If singleton meets vacancy: Active nodes update state
- Singletons break off with probability ϵ



Critical case: #Atoms = Integer multiple of final assembly

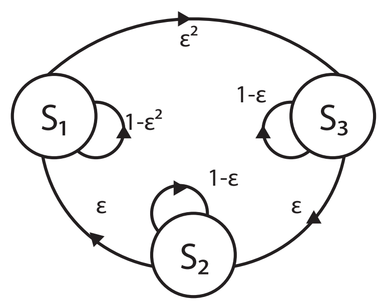


- **Fox & JSS (2009):** A state is stochastically stable if and only if there is a minimal number of (sub)assemblies.
- Corollary: Let a complete assembly have N parts. The maximum number of incomplete assemblies is $N-1$. (For any number of atoms.)

- *Stochastic stability* definition:
  - Let P^ϵ denote the transition probability matrix of an irreducible & aperiodic Markov chain.
  - Let μ^ϵ be the (unique) stationary distribution for P^ϵ
  - A state, x, is **stochastically stable** if

$$\liminf_{\epsilon \to 0} \mu^\epsilon(x) > 0$$

- Trivial illustration:



- **Young (1993):** Stochastic stability via resistance trees.

- A "language" $\mathcal{L}$ is a pair of matrices (P, Q)
  - Binary elements, row sum = 1
  - Speaker matrix: P : events $\rightarrow$ words
  - Hearer matrix: Q : words $\rightarrow$ events
- Illustration:

$$P = \begin{array}{cccc} & \alpha & \beta & \gamma \\ A & 1 & 0 & 0 \\ B & 1 & 0 & 0 \\ C & 0 & 1 & 0 \end{array} \qquad Q = \begin{array}{cccc} & A & B & C \\ \alpha & 1 & 0 & 0 \\ \beta & 0 & 1 & 0 \\ \gamma & 0 & 0 & 1 \end{array}$$

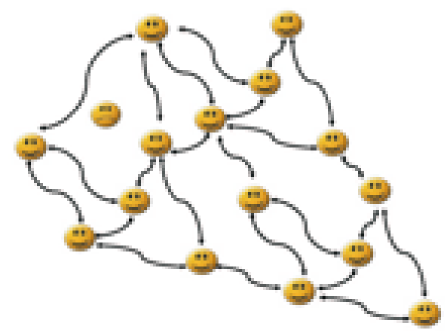
- Optimal language: maximum tr$[PQ]$ or $P = Q^{\mathsf{T}}$
- Assume square matrices for convenience
- Population of agents, $\mathcal{I} = \{1, ..., \ell\}$
- Fitness of agent i with language $\mathcal{L}_i = (P_i, Q_i)$:

$$f_i = \mathrm{tr}[P_i \frac{1}{\ell} \sum_{k=1}^{\ell} Q_k] + \mathrm{tr}[\frac{1}{\ell} \sum_{k=1}^{\ell} P_k Q_i]$$

- Update rules:
  - Global:
    - Select agent i at random
    - Update:

$$\mathcal{L}_i^+ = \begin{cases} \arg\max_k f_k & \text{w.p. } 1-\epsilon \\ \text{rand} & \text{w.p. } \epsilon \end{cases}$$

  - Local:
    - Connected undirected graph
    - Select edge (i,j) at random
    - Update: Assuming $f_i \geq f_j$



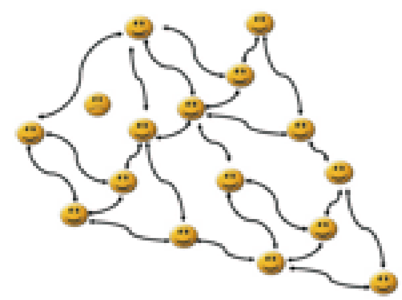
$$\mathcal{L}_j^+ = \begin{cases} f_i & \text{w.p. } 1-\epsilon \\ \text{rand} & \text{w.p. } \epsilon \end{cases}$$

- Unperturbed ($\epsilon = 0$) recurrence class: Consensus
- **Fox & JSS (2011):** A state is stochastically stable if and only if it is a uniform optimal language.
  Proof: Resistance tree arguments.

Final remarks

	Stability	Selection
Descriptive	*explanation*	*refinement*
Prescriptive	*adaptation*	*efficiency*

- Recap: Dynamics matter!
  - Main tools:
    - ∗ Stochastic approximation
    - ∗ Stochastic stability
  - Both prescriptive and descriptive agenda



- Absent: Convergence rates
  (cf., Saberi, Shah & coauthors)
- Future work:
  - "Natural" learning rules?
  - Fully exploit prescriptive agenda (e.g., chatter)
  - Agent states